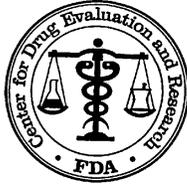


**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:
202020Orig1s000

STATISTICAL REVIEW(S)



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Science
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number: 202020

Drug Name: NP01 (prednisone, modified-release) tablets

Indication(s): Treatment of rheumatoid arthritis in adult patients

Applicant: Horizon Pharma, Inc.

Date(s): Receipt date: September 26, 2011
PDUFA date: July 26, 2012

Review Priority: Standard

Biometrics Division: Division of Biometrics II

Statistical Reviewer: Kiya Hamilton, Ph.D.

Concurring Reviewers: Ruthanna Davi, Ph.D., Statistical Reviewer

Medical Division: Division of Pulmonary, Allergy and Rheumatology Products

Clinical Team: Rosemarie Neuner, M.D., Medical Reviewer
Susan Limb, M.D., Team Leader
Badrul A. Chowdhury, M.D. Ph.D., Medical Division Director

Project Manager: Michelle Jordan Garner

Keywords: NDA review, clinical studies

Table of Contents

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES	1
FOOD AND DRUG ADMINISTRATION	1
STATISTICAL REVIEW AND EVALUATION	1
LIST OF TABLES.....	3
LIST OF FIGURES.....	4
1. EXECUTIVE SUMMARY	5
2. INTRODUCTION	5
2.1 OVERVIEW.....	5
2.2 DATA SOURCES	7
3. STATISTICAL EVALUATION	7
3.1 DATA AND ANALYSIS QUALITY	7
3.2 EVALUATION OF EFFICACY	7
3.3 EVALUATION OF SAFETY	18
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	18
4.1 GENDER, RACE AND AGE	18
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS	19
5. SUMMARY AND CONCLUSIONS	20
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE	20
5.2 COMMENTS ON THE PROPOSED LABEL	21
5.3 CONCLUSIONS AND RECOMMENDATIONS	21

LIST OF TABLES

Table 1 Summary of Study Design.....	8
Table 2 Summary of Patient Disposition.....	12
Table 3 Demographics and Baseline Characteristics (mITT population).....	13
Table 4 Summary of Analysis Methods	13
Table 5 Analysis of the ACR20 Response at Visit 4 Using Different Imputation Schemes (mITT Population)	16
Table 6 Analysis of the Relative Change from Baseline in the Duration of Morning Stiffness at Visit 4- Key Secondary Variable (mITT Population, LOCF, BOCF).....	17
Table 7 Additional Analysis of the Relative Change from Baseline in the Duration of Morning Stiffness at Visit 4- Key Secondary Variable (full mITT Population, LOCF, BOCF).....	18
Table 8 Subgroup Analysis of ACR20 Response at Visit 4 by Sex and Age: mITT Population (Worse Case Imputation)	19
Table 9 Subgroup Analysis of ACR20 Response at Visit 4 by Region and Disease Duration: mITT Population (Worse Case Imputation).....	20

LIST OF FIGURES

Figure 1 Study Design.....	9
----------------------------	---

1. EXECUTIVE SUMMARY

Horizon Pharma proposes a delayed-release (DR) formulation of prednisone, RAYOS[®] (NP01), for the treatment of rheumatoid arthritis in adult patients. The applicant conducted Study NP01-007 with the primary efficacy objective of evaluating if 12 weeks of treatment with 5 mg DR RAYOS[®] (NP01) administered in the evening was superior to placebo in terms of the percentage of patients achieving American College of Rheumatology (ACR) 20 response.

Based on Study NP01-007, there is substantial evidence of efficacy of RAYOS[®] (NP01). The difference in proportions for ACR20 at visit 4 is 17 with a p-value of 0.001. The relative change from baseline in duration of morning stiffness, an endpoint not necessary for approval but defined a priori as a key secondary endpoint, has a difference in medians of -19 between the two treatment groups with a p-value of 0.0015 (last observation carried forward). The result for morning stiffness is not replicated in another study, as the sponsor was told would be necessary for description of this endpoint in the Clinical Studies section of labeling. Statistical issues regarding the analysis population (modified Intention-to-Treat), missing data handling, subpopulations based on baseline demographics, and choice of statistical models with covariates were considered as part of the review and ultimately were sufficiently resolved to conclude that the efficacy of RAYOS[®] (NP01) has been demonstrated.

In the primary analysis of ACR20 response rates, missing values were imputed as non-responders, and is referred to as the worse case imputation. The applicant also conducted sensitivity analyses using other missing data imputations such as last observation carried forward (LOCF), withdrawal imputation and observed case. In the analysis of the key secondary endpoint, duration of morning stiffness between baseline and visit 4, the applicant applied LOCF and baseline observation carried forward (BOCF) methods. There was a 6% dropout rate for the RAYOS[®] (NP01) group and 11% for the placebo group. The most common reason for early dropout in both treatment groups was due to adverse events (AE); 6 (3%) patients in the RAYOS[®] (NP01) group and 8 (7%) patients in the placebo group. Given that the findings were consistent after applying different imputation strategies, the impact of missing data in this application is considered inconsequential. There were no adjustments for multiplicity.

The analysis of the key secondary endpoint provided by the sponsor did not include the entire intent-to-treat group. Initially, no explanation was provided for this discrepancy in the sample size. At the FDA's request the applicant conducted an additional analysis on this endpoint for the full modified-intent-to-treat (mITT) population and explained that in the original analysis, patients with missing baseline values had been excluded. In the additional analysis, if the patient had a missing baseline value, the applicant used the screening value then LOCF was applied. The qualitative conclusions from the supplemental analysis were consistent with those of the analysis in the original submission.

2. INTRODUCTION

2.1 Overview

2.1.1 Class and Indication

Horizon Pharma, Inc. proposes RAYOS[®] (NP01), a delayed-release (DR) prednisone tablet, administered once daily in the evening at approximately 10 p.m. for the treatment of rheumatoid arthritis (RA). RAYOS[®] (NP01) will be referred to as NP01 for the duration of this review.

RA is an autoimmune disease that causes chronic inflammation of the tissues around the joints, as well as in other organs in the body. This chronic, symptomatic disease can result in a variety of outcomes with different chronologies, severities, and overall patient effects. The applicant stated the following:

Disease-related symptoms in patients with chronic inflammatory disease exhibit a pronounced circadian rhythm. Clinical symptoms of RA such as pain and morning stiffness of the joints are more severe early in the morning after awakening than in the afternoon or evening. Historically, morning stiffness, being a common typical symptom of RA, was a standard diagnostic criterion for the disease.

Morning stiffness has been shown to be one of the most important clinically relevant symptoms of RA.

Glucocorticoids have been used in the treatment of RA since 1948, mainly because of their ability to relieve symptoms such as joint stiffness and joint pain but also because they slow down disease progression. The glucocorticoids most widely used today, such as predniso(lo)ne, have a short half-life of 2 to 3 hours and are usually given in the morning at a low oral dose in order to minimize the disturbance of the physiological control of the endogenous adrenal steroid release cycle. The concept of low-dose corticoid therapy in the treatment of RA is well established, and its safety and effectiveness have been confirmed by several clinical trials.

Horizon Pharma is requesting approval for dosage strength of 5 mg once daily of DR NP01 that is to be administered in the evening. The new formulation will release prednisone approximately four hours after being taken by the patient. The applicant claims that the pharmacokinetic behavior is identical to that of standard immediate-release (IR) prednisone. According to the applicant,

NP01 represents a novel prednisone formulation that allows low-dose chronotherapy of RA. By synchronizing the prednisone release to the circadian rhythms of inflammation, NP01 provides better efficacy compared to the same prednisone dose level administered as an IR formulation. Since the safety profile is unchanged, this result may be an improved risk-benefit ratio for low-dose prednisone.

2.1.2 History of Drug Development

A pre-IND meeting was held on March 26, 2006.

(b) (4)

In addition, the Division stated that a representative American population would need to be investigated or a rationale would need to be provided for why these investigations are not needed in the context of a rheumatoid arthritis population. An end-of-phase 2 meeting was held on December 13, 2007. During this discussion, the Division

indicated that if successful, study NP01-007 alone could support an indication for signs and symptoms of rheumatoid arthritis. The Division also indicated that a labeling claim for improvement in morning stiffness would need to be supported by two studies and would be described only in the Clinical Studies section, not the Indication section. The Division made the following additional comments regarding a nonresponder analysis for ACR20.

In the protocol synopsis (# NP01-007), you propose to employ the last observation carried forward (LOCF) approach to impute missing data for the primary endpoint (i.e. ACR20 responder). The Division recommends that any subject who drops out of the study be considered a non-responder.

A pre-NDA meeting was held on January 26, 2010. The Division reiterated that morning stiffness would be described in the Clinical Studies section only, not in the Indication section. In response to the applicant's question regarding the key secondary endpoint, reduction of morning stiffness, the division stated that

Your LOCF imputation strategy for patients that discontinued early is not appropriate. You may be imputing a good score for a patient that was not able to tolerate study drug. You should use baseline observation carried forward for these patients in the primary analysis, i.e., there would be no improvement in the baseline score. You may explore other imputation strategies as part of your sensitivity analyses.

2.1.3 Specific Studies Reviewed

Efficacy Study NP01-007 is the focus of this review. Study NP01-007 is a phase 3, randomized, double-blind, placebo-controlled, parallel group, multi-center 12 week study in patients 18 to 80 years of age with RA. Approximately 350 patients were randomized to NP01 5 mg or placebo in a 2:1 ratio (231 in the NP01 group, 119 in the placebo group). There were a total of 62 study centers in North America (United States and Canada) and Europe (Germany, Hungary, Poland and the United Kingdom).

2.2 Data Sources

All data was supplied by the applicant to the CDER electronic data room in SAS transport format. The data and final study report for the electronic submission were archived under the network path location <\\CDSESUB1\EVSPROD\NDA202020\202020.enx>. The information needed for this review was contained in modules 1, 2.5, 2.7 and 5.3.5.

3. STATISTICAL EVALUATION

3.1 Data and Analysis Quality

In general, the submitted efficacy data are acceptable in terms of quality and integrity. The primary and secondary efficacy endpoint analyses are reproducible.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

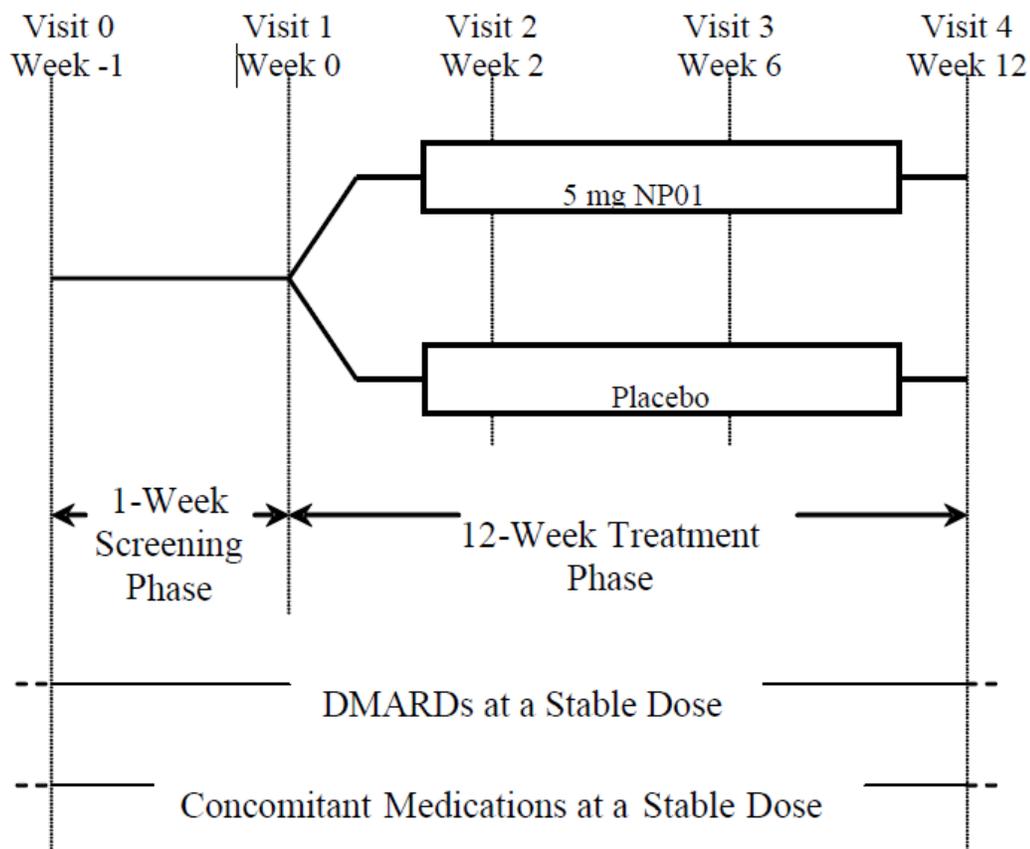
The summary of the study design and endpoints are given in Table 1. Study NP01-007 was a phase 3, randomized, double-blind, placebo-controlled, parallel-group, multi-center 12 week study with an initial 1-week screening phase in male and female subjects 18 to 80 years of age who were receiving disease-modifying anti-rheumatic (DMARD) treatment. The design and efficacy endpoints are explained in detail in the following paragraphs.

Table 1 Summary of Study Design.

Study ID	Indication	Length of the Study	Treatment Arms	Number of Patients	Primary Efficacy Endpoints
NP01-007	RA	SP: 1 week TP: 12 weeks	NP01, 1 tablet 5 mg Placebo, 1 tablet 0 mg	231 119	ACR20 responder status at Visit 4 (Week 12)

- SP: Screening Phase
- TP: Treatment Phase

Study NP01-007 was designed to assess the efficacy and safety of evening administration (at approximately 10 p.m.) of NP01 in comparison to placebo in patients with RA. The study consisted of a 1 week single blinded screening phase, visit 0 to visit 1 (week -1 to week 0), and a 12 week treatment phase; which consisted of visit 1 (baseline; week 0), visit 2 (week 2), visit 3 (week 6) and visit 4 (week 12). Patients remained on their RA medication during both the screening phase and the treatment phase. The following diagram depicts a schematic of the study design.



DMARD = disease modifying anti-rheumatic drug

Figure 1 Study Design

Excerpted from the Clinical Study Report, page 31 Figure 1

In the screening phase patients received placebo in addition to their standard RA medication.

In the 1-week screening phase, patients made daily recordings in a patient diary about their morning stiffness, pain intensity, intake of study medication and use of analgesics. Entries regarding duration of morning stiffness prior to Visit 1 were used to calculate a robust baseline value for this variable (average of 7 daily values collected on Days -7 to -1). The purpose of the 1-week screening phase was to establish whether the patient’s compliance with study medication and completion of study diaries were sufficient to qualify a screened patient for randomization to study treatment at Visit 1.

Following the screening phase, patients were randomized in a 2:1 ratio to NP01 (5 mg/day) or placebo.

Patients continued to make daily recordings in a patient diary about their morning stiffness, pain intensity, intake of study medication, and use of analgesics. Rheumatoid disease status, concomitant medications, vital signs, and AEs were assessed at all visits.

After the double-blind treatment phase, patients were to be switched to 5 mg IR predniso(lo)ne and tapered down according to best practice, if applicable.

The sponsor noted that

The risk-benefit ratio of the study design was considered favorable because (i) all patients received standard DMARD therapy for their RA throughout the study, (ii) a 2:1 randomization was chosen to minimize the number of patients receiving placebo treatment, and (iii) patients with deterioration of their disease were to be withdrawn from the study.

The primary efficacy endpoint was the ACR20 response rate at visit 4 (week 12), which was defined as the proportion of subjects meeting the ACR criteria of at least 20% improvement in tender and swollen joint counts and at least a 20% improvement in 3 of the remaining 5 core set measures (patient's assessment of pain (visual analog scale [VAS]), patient's global assessment of disease activity (VAS), physician's global assessment of disease activity (VAS), functional disability index of the health assessment questionnaire (HAQ-DI), and one acute phase reactant value [C-reactive protein (CRP) or erythrocyte sedimentation rate (ESR)]). Note the CRP were used; the ESR results were used if the CRP result were not available.

The key secondary efficacy variable was the relative change (i.e., percent change from baseline) in the duration of morning stiffness between baseline and visit 4 (week 12). Duration of morning stiffness was defined as the difference between the time of resolution of morning stiffness and the time of wake-up.

The latest resolution time for morning stiffness was set to noon; therefore, if the time when the morning stiffness eased was after noon, this was censored at noon, and duration of morning stiffness (min) was noon minus the time of getting up in the morning. Duration of morning stiffness was calculated as the average of the morning stiffness duration over the last 7 days prior to the visit (including the day of the visit). If more than 4 assessments were missing the duration was set to missing. If at least 4 assessments were available, the 7-day average was computed using the last 7 days with available values prior to the visit day for the duration of morning stiffness.

The other secondary efficacy variables were as follows:

- ACR20 response rate at Week 2 (Visit 2) and Week 6 (Visit 3).
- ACR50 response rate at Week 2 (Visit 2), Week 6 (Visit 3), and Week 12 (Visit 4); responders were defined as per ACR20 but using 50% reduction instead of 20%.
- ACR70 response rate at Week 2 (Visit 2), Week 6 (Visit 3), and Week 12 (Visit 4); responders were defined as per ACR20 but using 70% reduction instead of 20%.
- Time to response based on ACR20 criteria.
- Change from baseline in disease activity score 28 (DAS28) at each visit.
- European League Against Rheumatism (EULAR) response criteria: patients were classified as patients with good, moderate, or no response based on their change in DAS28.
- Relative (%) and absolute reduction of duration of morning stiffness between baseline and each study visit.
- Change from baseline in severity of morning stiffness at each visit (100 mm VAS).
- Change from baseline in terms of reoccurrence of stiffness during day (while performing routine activities). Reoccurrence of stiffness during the day was assessed as the percentage of days with reoccurrence of stiffness over the last 7 days prior to each visit (if 4 or more responses were missing, the percentage was set to missing).
- Change from baseline in tender and swollen joint counts at post-baseline study visits. The analysis of tender joint count and swollen joint count was based on a 28-joint assessment. For

- each patient, only those joints that were evaluable at baseline and endpoint were included in the statistical analysis of joint counts.
- Change from baseline in patient assessment of the pain intensity at each visit (100 mm VAS).
 - Change from baseline in physician's and patient's global assessments of disease activity at each visit (100 mm VAS).
 - Change from baseline in HAQ-DI at each visit: the maximum score of all items within each of the 8 categories determined the category score for each patient. The HAQ-DI was the average of all 8 category scores.
 - Change from baseline in inflammatory parameters at each visit (CRP, ESR, tumor necrosis factor [TNF α], and interleukin-6 [IL-6])
 - Change from baseline in urine C-terminal cross-linked telopeptides of collagen type I (CTX I).
 - Change from baseline in occurrence of pain in the morning and evening (100 mm VAS) was computed as the change in percentage of occurrence of pain in morning/evening over the last 7 days prior to each visit (if 4 or more responses were missing, the percentage was set to missing).
 - Change from baseline in the use of additional analgesics (yes/no) was computed as the change from baseline in percentage of days with the event over the last 7 days prior to each visit (if 4 or more responses were missing, the percentage was set to missing).
 - Use of additional analgesics was assessed as the number of days with additional analgesics during the double-blind treatment phase.
 - Change from baseline in each domain of the Quality of Life (QoL) questionnaire short form (SF-36) and for the mental and physical component scores.
 - Change from baseline in the fatigue subset of the Functional Assessment of Chronic Illness Therapy- Fatigue (FACIT-F) questionnaire.

ACR 50 and ACR 70 responses are defined similarly to ACR 20 with 50% and 70% improvement, respectively.

3.2.2 Patient Disposition, Demographic and Baseline Characteristics

The applicant's target population was patients with RA 18 to 80 years of age. All patients had a documented history of RA, in agreement with the ACR criteria, including the symptoms morning stiffness, joint pain, tender and swollen joints, and inflammatory state with elevated ESR or CRP.

A summary of the patient disposition in Study NP01-007 is given in Table 2. There were 231 and 119 patients randomized to NP01 and placebo, respectively. The mITT population was defined as all randomized patients who received at least 1 dose of study medication and therefore since all randomized subjects received at least one dose of study medication, the mITT population included 231 and 119 patients in the NP01 and placebo groups, respectively.

The safety population included all randomized patients who received at least one dose of study medication and were analyzed according to the treatment which they actually received. The per-protocol (PP) population was a subset of the mITT population and included all patients who did not have a major protocol deviation. The relevant major protocol deviations are as follows:

Inclusion/Exclusion/Randomization Criteria:

- Patients with an average duration of morning stiffness of <35 minutes at baseline (Visit 1).
- Tender/Swollen joint count at baseline (Visit 1) <4.

- Patients who had not taken disease modifying anti-rheumatic drugs (DMARDs) for >6 months prior to screening (Visit 0) or who had stopped DMARDs during the study.

Misrandomization:

- Patients who were misrandomized and were on a different treatment after unblinding.

Compliance/Exposure:

- Overall exposure <75 days for completers.
- Overall study drug compliance (double-blind phase) <80%.
- Patients who had taken the drug outside of 10 p.m. ±1 hour at least once out of the 7 days leading up to visit and no more than 3 times outside of 10 p.m. ± 1 hour throughout the entire diary entry period

Prohibited Medication

- Patients who took systemic glucocorticoids that stopped <4 weeks before the screening date (intake after last dose of study medication was not considered a protocol violation).
- Patients who had unstable non-steroidal anti-inflammatory drugs (NSAIDs) (intermittent use >7 days) or patients who had started taking oral NSAIDs after the double-blind phase which was not due to AEs. (NSAIDs that stopped during the study and started before were not considered protocol deviations).
- Intake of biologicals (case-by-case review).

The most common reason for early termination was patient request, about 3% for the NP01 group and about 7% for the placebo group.

Table 2 Summary of Patient Disposition

	NP01	Placebo
Randomized	231	119
mITT population	231	119
Safety population	231	119
PP Population	185	94
Discontinued	14 (6%)	13 (11%)
Patient request	6	8
Adverse Event	5	2
Lost to Follow-up	1	1
Other reasons	2	2

Selected demographic and baseline characteristics are summarized in Table 3 for the mITT population. The patients' mean age was about 57 years. The majority of the patients were women (84%), Caucasian (98%) and with Ethnicity not assessed (European, Canada) (76%). There were no noticeable imbalances of the demographics and baseline characteristics between the treatment groups on gender, age, race and ethnicity.

Table 3 Demographics and Baseline Characteristics (mITT population)

		NP01 N=231	Placebo N=119
Age (years)	Mean (SD)	57.1 (9.89)	57.5 (9.55)
Age category, [n (%)]	Young (≤45 yrs)	24(10.4%)	12 (10.1%)
	Middle-aged (>45 to ≤65 yrs)	162 (70.1%)	83 (69.7%)
	Elderly (>65 to ≤75 yrs)	35 (15.2%)	20 (16.8%)
	Very elderly (>75 yrs)	10 (4.3%)	4 (3.4%)
Gender [n(%)]	Female	192 (83.1%)	102 (85.7%)
	Male	39 (16.9%)	17(14.3%)
Ethnicity [n (%)]	Not Hispanic/Latino	49 (21.2%)	23 (19.3%)
	Hispanic/Latino	6 (2.6%)	5 (4.2%)
	Not assessed (EU, Canada)	175 (75.8%)	90 (75.6%)
	Missing	1 (0.4%)	1 (0.8%)
Race [(n<100)]	White	226 (97.8%)	118 (99.2%)
	Black or African American	4 (1.7%)	1 (0.8%)

Source: Study NP01-007, Clinical Study Report Section 11.2.1 Table 11.

The primary and secondary efficacy analyses were conducted on the mITT population according to the treatment to which they were randomized. The applicant also conducted the primary and secondary analyses on the safety population and the analysis of the primary efficacy endpoint, ACR20, was performed on the per-protocol (PP) population.

3.2.3 Statistical Methodologies

The primary objective of Study NP01-007 was to evaluate if 12 weeks of treatment with 5 mg DR NP01 administered in the evening was superior to placebo in terms of the percentage of patients achieving ACR20 response. The summary of the statistical methods used for the primary and key secondary analysis is given in Table 4 followed by further description of the analyses in the subsequent text.

Table 4 Summary of Analysis Methods

Study ID	Measure of Interest	Analysis Method	Adjustments
NP01-007	ACR20	Logistic regression	No multiplicity adjustments
	Morning Stiffness	Median and 95% CI of the median using the Hodges Lehmann method	No multiplicity adjustments

The primary efficacy endpoint, ACR20 response at visit 4 (week 12), was analyzed using a logistic regression model with treatment (NP01 5 mg/day versus placebo), geographic region, median age class and gender as covariates. The sponsor also conducted the following analyses on

the primary efficacy endpoint presumably to address, at least in part, the concern regarding geographical region expressed by the Division at the pre-IND meeting:

In order to evaluate the consistency of results across sites, treatment p-values for the model with no interaction term and models with each of the 2 types of interaction terms added into the logistic regression model with treatment, geographic region, gender, and median age class as factors were presented. The 2 types of interactions were as follows:

- An interaction between the treatment and the nested effect, where the nested effect was the pooled sites within the geographic region (defined as US/Canada and Europe).
- An interaction between the treatment and region, where region was defined as US/Canada versus Europe.

Race was not included in the analyses of the primary efficacy variable because 98% of the patients were White. The analyses of ACR20 were repeated for the Safety and PP populations as sensitivity analyses. Odds ratios (NP01: placebo) for the difference between treatments and the associated 95% confidence intervals (CI), as well as the difference in proportion and its associated 95% CI were presented for each population.

According to the sponsor, the key secondary efficacy endpoint (the relative change (%) in the duration of morning stiffness) was not expected to be normally distributed. Thus, the Hodges Lehmann method was used to estimate the difference in medians and the associated 95% confidence interval (CI) to be used for inferential comparison of the treatment groups. The Wilcoxon rank sum test was used to compute the p-value for the comparison between treatment groups; however, this was provided for informational purposes only.

Secondary efficacy endpoints, ACR50 and ACR70, were tested separately using a logistic regression model with treatment as a factor and a 2-sided significance level of 0.05. DAS28 was analyzed using a mixed analysis of covariance (ANCOVA) model with baseline, treatment, pooled sites as a nested effect of geographic region (random effect) and the interaction between the nested effect and the treatment. Time to response was analyzed using the Kaplan-Meier method and treatments were compared using a Cox model stratified by geographic region. EULAR response was analyzed using logistic regression with treatment and geographic region as factors.

For the primary efficacy endpoint, ACR20 responder rate at visit 4, all patients with missing data at visit 4 were imputed as non responders, and is referred to as the worse case imputation method. The applicant also conducted sensitivity analyses including an observed case imputation (i.e., no imputation for missing values performed), last observation carried forward (LOCF) and withdrawal imputation. Withdrawal imputation was defined as follows, if

A patient discontinued prematurely, the ACR20 was imputed as non-responder, while patients who completed the study but had a missing assessment for ACR20 at Visit 4 were not imputed.

For the key secondary endpoint, change in duration of morning stiffness, missing values were imputed using LOCF and baseline observation carried forward (BOCF). LOCF was the primary missing data imputation method used contrary to the Division's advice to use BOCF as the primary imputation method.

The applicant conducted subgroup analyses for the primary efficacy endpoint based on region (Canada/UAS, Europe), sex (female, male), age group (<40 years, 40 to <65 years, ≥65 years) and disease duration (≤5.94 years and >5.94 years).

3.2.4 Results and Conclusions

Table 5 below shows the primary efficacy analysis results for ACR20 responder rate at visit 4 (week 12) for the mITT population using worse case imputation method. The responder rate at visit 4 was greater in the NP01 group than in the placebo group at 47 % and 29%, respectively. This was a significant difference in proportions between the treatment groups of 17% (95% CI: 7, 28), p-value=0.001. The results from the sensitivity analyses using observed case, LOCF and withdrawal imputations are also included in the Table 5. The qualitative conclusions regarding the treatment effect are the same for all three imputation strategies for the sensitivity analyses as for the worse case imputation method. The applicant also conducted the primary analysis on the PP population. The same conclusion was drawn from the PP analysis as from the mITT analysis.

Table 5 Analysis of the ACR20 Response at Visit 4 Using Different Imputation Schemes (mITT Population)

Imputation scheme	NP01 n/N (%)	Placebo n/N (%)	% Difference in proportions ^a (95% CI) ^b		Odds Ratio (95% CI) ^c	P-value ^d
Primary analysis						
Worse case ^e	108/231 (47%)	34/119 (29%)	18	17 (7, 28)	2 (1, 4)	0.001
Secondary analysis						
Observed case	108/224 (48%)	35/116 (30%)	18	17 (7, 28)	2 (1, 4)	0.0013
LOCF ^f	110/229 (48%)	35/119 (29%)	18	18 (8, 28)	2 (1, 4)	0.0007
Withdrawal ^g	108/230 (47%)	35/119 (29%)	18	17 (7, 27)	2 (1, 4)	0.0014

Source: Study NP01-007, Clinical Study Report Section 11.4.1.1 Table 18.

CI = confidence interval; N = total number of patients per treatment group and imputation scheme at corresponding visit; n = number of responders.

Note: Visit 4 includes early withdrawal patients.

^a The observed difference between treatments (first value) and the estimate of the treatment difference from the generalized linear model (second value) are reported.

^b The 95% CI was calculated from a generalized linear model with a binomial probability function and an identity link with treatment, geographic region, gender, and median age class as factors.

^c Asymptotic 95% CIs based on asymptotic normality of the estimated odds ratio.

^d The p-value was based on logistic regression with treatment, geographic region, gender, and median age class as factors.

^e Worse case imputation: all missing values were imputed as non-responders.

^f LOCF imputation: last observation (post-baseline) was carried forward.

^g Withdrawal imputation: missing values for withdrawn patients were imputed as non-responders.

The analysis of the key secondary efficacy endpoint provided by the sponsor did not include the entire mITT group. Initially, no explanation was provided for this discrepancy in the sample size. At the FDA's request, the applicant conducted an additional analysis of the relative change from baseline in duration of morning stiffness for the full mITT population. The applicant stated:

The sample size discrepancies can be explained by two key definitions in the statistical analysis plan (SAP):

- mITT population: all patients who were randomized and received at least 1 dose of study medication. Patients are analyzed according to the treatment to which they were intended to be randomized.
- Assessments of diary data will be used on the last 7 days prior to each visit. If more than 4 assessments during this 7 day period are missing, the assessment will be set to missing. In this case only, a LOCF method will be applied for imputing missing assessments and will consist of taking the last 7 nonmissing entries of the diary data prior to the visit. If strictly less than 3 assessments are available using LOCF, the value will be set to missing.

The mITT definition does not require patients to have either a baseline or a post-baseline measurement present. In addition, morning stiffness was collected in the diary and LOCF was carried out in a different manner for diary data than for the other collected data. Essentially the LOCF for diary data just looked back to the previous week in the diary and did not go all the way

back to the previous visit and carry forward. Thus whenever a patient dropped out of the study or did not fill out the diary during the critical period when data was considered evaluable then that patient was not analyzed/counted in N for the subsequent visit(s).

Some patients were missing the baseline value for the key secondary efficacy endpoint for the mITT population. This led to a reduction in the sample size for the analysis of this endpoint. The sample size was reduced from 231 in the NP01 group and 119 in the placebo group (full mITT population) to 216 in the NP01 group, 107 in the placebo group for the LOCF analysis and 215 in the NP01 group, 107 in the placebo group for the BOCF analysis.

In the additional analysis, if the baseline value was missing the applicant replaced it with the patients screening value and LOCF and BOCF imputations were applied. The applicant noted that

There is one patient in NP01 treatment group reported as having duration of morning stiffness of 0 minutes at baseline/screening and was captured as a protocol deviation. Due to this 0 value, the relative change cannot be calculated for this patient, therefore the N for the NP01 group is reduced by 1 (N=230) for the LOCF analysis of relative change.

Table 6 shows the results for the relative change in the duration of morning stiffness between baseline and visit 4 for the mITT population. Patients with missing baseline values were excluded in this analysis. Table 6 includes the outcomes for the LOCF and BOCF missing data imputation schemes. For both the LOCF and the BOCF, the median relative change from baseline was greater in the NP01 group compared to the placebo group. The difference in the median relative change was -20% (95% CI: -32, -6) for the LOCF and -20% (95% CI: -32, -7) for the BOCF.

Table 6 Analysis of the Relative Change from Baseline in the Duration of Morning Stiffness at Visit 4- Key Secondary Variable (mITT Population, LOCF, BOCF)

Imputation scheme	Relative Change (%)				Difference in median ^a [%] (95% CI) ^a	P-value ^b
	NP01		Placebo			
	N	Median	N	Median		
LOCF	216	-55	107	-34	-20 (-32, -6)	0.0015
BOCF	215	-55	107	-33	-20 (-32, -7)	0.0013

Source: Study NP01-007, Clinical Study Report Section 11.4.1.2.1 Table 22.

CI = confidence interval; LOCF = last observation carried forward; N = number in analysis (i.e., excluding the missing).

Note: Visit 4 includes early withdrawal patients.

^a Difference in median and its 95% CI were estimated using Hodges-Lehmann method.

^b Wilcoxon signed rank test p-value (for information only).

As different methods were applied to compute the p-value and the 95% CI, the difference between treatment groups was assessed using the 95% CI.

Note: LOCF and BOCF imputation algorithms were implemented up to the next visit following the last diary data. LOCF (last observation carried forward) was computed using the last 7 days prior to the visit day with non-missing values for the duration of morning stiffness.

BOCF: Baseline observation was carried forward.

The results from the additional analysis for the key secondary efficacy endpoint are given in Table 7 below. The table includes the outcomes for the LOCF and BOCF missing data

imputation schemes. For both the LOCF and the BOCF, the median relative change from baseline was greater in the NP01 group compared to the placebo group. The difference in the median relative change was -20% (95% CI: -32, -7) for the LOCF and -18% (95% CI: -31, -6) for the BOCF.

Table 7 Additional Analysis of the Relative Change from Baseline in the Duration of Morning Stiffness at Visit 4- Key Secondary Variable (full mITT Population, LOCF, BOCF)

Imputation scheme	Relative Change (%)				Difference in median ^a [%] (95% CI) ^a	P-value ^b
	NP01		Placebo			
	N	Median	N	Median		
LOCF	230	-54	119	-27	-20 (-32, -7)	0.0006
BOCF	231	-51	119	-25	-18 (-31, -6)	0.0011

Source: Response to Information Request dated April 4, 2012

CI = confidence interval; LOCF = last observation carried forward; N = number in analysis (i.e., excluding the missing).

Note: Visit 4 includes early withdrawal patients.

^a Difference in median and its 95% CI were estimated using Hodges-Lehmann method.

^b Wilcoxon signed rank test p-value (for information only).

As different methods were applied to compute the p-value and the 95% CI, the difference between treatment groups was assessed using the 95% CI.

Note: LOCF and BOCF imputation algorithms were implemented up to the next visit following the last diary data. LOCF (last observation carried forward) was computed using the last 7 days prior to the visit day with non-missing values for the duration of morning stiffness.

BOCF: Baseline observation was carried forward.

The results for the secondary efficacy endpoint, ACR50, were consistent with ACR20. ACR70 showed a greater increase in the NP01 group compared to the placebo, however, the difference was not statistically significant for any of the missing data imputations. DAS28 and EULAR both had a greater increase in the NP01 group compared to the placebo group.

3.3 Evaluation of Safety

The evaluation of safety was conducted by Dr. Rosemarie Neuner. Reader is referred to Dr. Rosemarie Neuner's review for this section.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

The applicant conducted a subgroup analysis on the primary efficacy endpoint (ACR20 response at visit 4) by region (Canada/USA, Europe), sex, age group (<40 years, 40 to <65 years, ≥65 years) and disease duration (≤5.94 years and >5.94). The subgroup analyses were performed using the mITT population. Analyses using the worse case missing data imputation are shown in this section.

4.1 Gender, Race and Age

Table 8 summarizes the subgroup analysis by gender and age for Study NP01-007. There was a higher ACR20 response rate in the NP01 group than in the placebo group for both males and females and age group 40-<65 years. The difference between the two groups was only

statistically significant for females and age group 40-<65 years. The subgroups that did not have a statistically significant difference between the two treatment groups had small sample sizes, i.e. subgroups males, and age groups < 40 years and ≥65 years and therefore these results are not considered an indication that NP01 is ineffective in these subgroups. Subgroup analysis by race was not conducted as approximately 98% of the patients were Caucasian.

Table 8 Subgroup Analysis of ACR20 Response at Visit 4 by Sex and Age: mITT Population (Worse Case Imputation)

	NP01 5 mg N=230 n/N (%)	Placebo N=119 n/N (%)	Difference in Proportions (%) (95% CI) ¹	Odds Ratio (95% CI) ²	Treatment P-value ²
Subgroup					
Sex					
Male	14/39 (36)	5/17 (29)	-4 (-36, 27)	1 (0.3, 4)	0.8714
Female	94/192 (49)	29/102 (28)	21 (10, 32)	3 (1, 4)	0.0006
Age					
<40 years	6/10 (60)	2/5 (40)		2 (0.2, 15)	0.7466
40 to <65 years	79/166 (48)	19/84 (23)	22 (11, 34)	3 (2, 6)	0.0002
≥65 years	23/55 (42)	13/30 (43)	0.2 (-22, 22)	1 (0.4, 3)	0.9766

Source: Study NP01-007, Clinical Study Report Section 11.4.2.8 Table 2.1.2.1.1.

Note: Visit 4 includes early withdrawal patients.

Worse Case Imputation: All missing values are imputed as non-responders.

n = number of responders.

N = total number of patients per treatment group and imputation scheme at corresponding visit.

¹ Difference in Proportions and its Confidence Interval are calculated from a generalized linear model with the binomial probability function as its response distribution function and the identity link as its link function.

Treatment, geographic region, gender, and median age class are fit in the model as factors.

² Odds Ratio, its Confidence Interval, and the treatment effect p-value are calculated from a logistic regression with treatment, geographic region, gender, and median age class as factors.

4.2 Other Special/Subgroup Populations

Table 9 summarizes the subgroup analysis by region and disease duration for Study NP01-007 for the mITT population using the worse case imputation scheme. There was a higher ACR20 response rate in the NP01 group than in the placebo group for Canada/USA and Europe and disease duration. The difference between the two groups was only statistically significant for Europe and where the disease duration was greater than 5.94 years. The subgroups that did not have a statistically significant difference between the two treatment groups had small sample sizes, i.e. subgroups Canada/USA and disease duration ≤5.94 years and therefore these results are not considered an indication that NP01 is ineffective in these subgroups.

Table 9 Subgroup Analysis of ACR20 Response at Visit 4 by Region and Disease Duration: mITT Population (Worse Case Imputation)

	NP01 5 mg N=230 n/N (%)	Placebo N=119 n/N (%)	Difference in Proportions (%) (95% CI) ¹	Odds Ratio (95% CI) ²	Treatment P-value ²
Subgroup					
Region					
Canada/USA	20/56 (36)	6/32 (19)	12 (-9, 33)	2 (1, 7)	0.1042
Europe	88/175 (50)	28/87 (32)	18.7 (6, 31)	2.18 (1, 4)	0.005
Disease Duration					
≤5.94 years	51/115 (44)	20/60 (33)	12 (-2, 27)	2 (1, 3)	0.1112
>5.94 years	57/116 (49)	14/59 (24)	22 (7, 37)	3 (2, 7)	0.0016

Source: Study NP01-007, Clinical Study Report Section 11.4.2.8 Table 2.1.2.1.1.

Note: Visit 4 includes early withdrawal patients.

Worse Case Imputation: All missing values are imputed as non-responders.

n = number of responders.

N = total number of patients per treatment group and imputation scheme at corresponding visit.

¹ Difference in Proportions and its Confidence Interval are calculated from a generalized linear model with the binomial probability function as its response distribution function and the identity link as its link function.

Treatment, geographic region, gender, and median age class are fit in the model as factors.

² Odds Ratio, its Confidence Interval, and the treatment effect p-value are calculated from a logistic regression with treatment, geographic region, gender, and median age class as factors.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

There were no statistical issues that would preclude approval identified during the course of this review. Statistical issues regarding the missing data handling and the mITT analysis population were examined as part of the review and ultimately were sufficiently resolved to conclude that the results of the efficacy analyses of NP01 in terms of ACR20 and morning stiffness remain reliable despite these issues.

- Of the 350 patients randomized, 6% dropped out in the NP01 group and about 11% dropped out in the placebo group. The most common reason for early termination in both groups was patient request; 6 (3%) in the NP01 group and 8 (7%). AEs was the second most common reason for early termination in both groups, 5 (2%) in the NP01 group and 2 (2%) in the placebo group. The applicant conducted sensitivity analyses and applied different imputation strategies. Given that the findings were consistent for each imputation strategy, the impact of missing data in this application is considered inconsequential.
- For the key secondary efficacy endpoint, at the FDA's request, the applicant provided an additional analysis of the relative change from baseline in duration of morning stiffness

for the full mITT population. The results were consistent with the applicant's initial analysis submitted with this NDA.

In addition, the choice of statistical models and covariates were assessed in the review of the primary and key secondary efficacy results. Finally, subgroup analyses based on baseline demographics were replicated and considered in the review of the primary and key secondary efficacy results.

5.2 Comments on the Proposed Label

The clinical team recommends the removal of the clinical studies section.

5.3 Conclusions and Recommendations

Based on the statistical review of Study NP01-007, there is evidence to support the claim of efficacy of DR NP01 5 mg per day in the treatment of patients 18 years of age or older with RA. There is evidence that the NP01 5 mg group has a higher responder rate compared to the placebo group for the primary endpoint, ACR20 responder rate after 12 weeks (visit 4) of treatment (worse case imputation method). The results were similar for different imputation methods, i.e., observed case, LOCF and withdrawal methods. There is evidence that the key secondary endpoint, change from baseline in duration of morning stiffness at visit 4, has a statistically significant difference between the two treatment groups in favor of NP01 5 mg. In order for morning stiffness to appear in the Clinical Trials section of the label, the sponsor was informed that morning stiffness needed to be replicated in a second study. The result for morning stiffness was not replicated in a second study. Based on the statistical review of Study NP01-007 and the ACR20 endpoint, there is substantial evidence to support the claim that NP01 5 mg is superior to placebo in patients with RA.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

KIYA HAMILTON
06/18/2012

RUTHANNA C DAVI
06/18/2012